Miscellaneous Paper D-94-4
June 1994

US Army Corps of Engineers
Waterways Experiment Station

Dredging Operations Technical Support Program

A Paradigm for Developing Sediment Toxicity Bioassays for the Regulatory Evaluation of Dredged Material

by Thomas M. Dillon

Prepared for Headquarters, U.S. Army Corps of Engineers

Dredging Operations Technical Support Program

Miscellaneous Paper D-94-4
June 1994

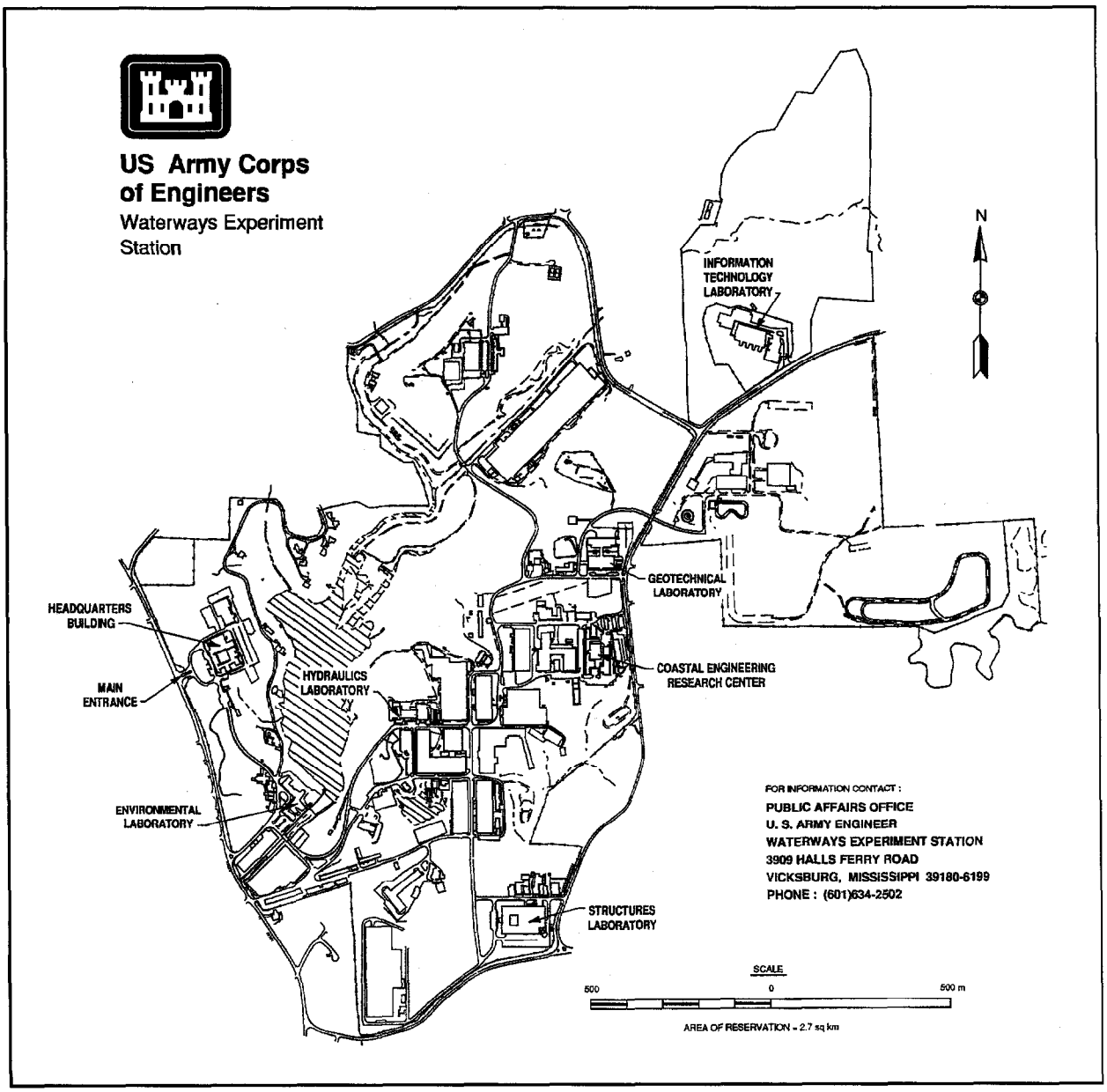
A Paradigm for Developing Sediment Toxicity Bioassays for the Regulatory Evaluation of Dredged Material

by Thomas M. Dillon

U.S. Army Corps of Engineers
Waterways Experiment Station
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Final report

Prepared for U.S. Army Corps of Engineers
Washington, DC 20314-1000



Waterways Experiment Station Cataloging-in-Publication Data

Dillon, Thomas M.
A paradigm for developing sediment toxicity bioassays for the regulatory evaluation of dredged material / by Thomas M. Dillon ; prepared for U.S. Army Corps of Engineers.
28 p. : ill. ; 28 cm. -- (Miscellaneous paper ; D-94-4)
Includes bibliographical references.
1. Dredging spoil -- Environmental aspects. 2. Biological assay. 3. Dredging spoil -- Law and legislation. 4. Spoil banks -- Testing. I. United States. Army. Corps of Engineers. II. U.S. Army Engineer Waterways Experiment Station. III. Dredging Operations Technical Support Program (U.S. Army Engineer Waterways Experiment Station. Environmental Laboratory) IV. Title. V. Series: Miscellaneous paper (U.S. Army Engineer Waterways Experiment Station) ; D-94-4.
TA7 W34m no.D-94-4

Contents

Preface

The work reported herein was conducted by the U.S. Army Engineer Waterways Experiment Station (WES) for Headquarters, U.S. Army Corps of Engineers (HQUSACE). Financial support was provided by HQUSACE through the Dredging Operations Technical Support (DOTS) Program, Task Area "Guidance for Determining Acceptability of Proposed Tests for the Open-water Disposal of Dredged Material." The DOTS Program is managed through the Environmental Effects of Dredging Programs, Dr. Robert M. Engler, Manager.

The report was prepared by Dr. Thomas M. Dillon, Fate and Effects Branch (FEB), Environmental Processes and Effects Division (EPED), Environmental Laboratory (EL), WES. The author gratefully acknowledges the input provided by numerous individuals during the course of this project. Especially helpful were Dr. Betsy Southerland, Mr. Tom Armitage, and Mr. Thomas Chase, U.S. Environmental Protection Agency, Office of Water.

The work was performed under the general supervision of Dr. Bobby L. Folsom, Jr., Chief, FEB, EPED. The Chief of EPED was Mr. Donald L. Robey, and Director of EL was Dr. John Harrison.

At the time of publication of this report, Director of WES was Dr. Robert W. Whalin. Commander was COL Bruce K. Howard, EN.

This report should be cited as follows:

> Dillon, T. M. (1994). "A paradigm for developing sediment toxicity bioassays for the regulatory evaluation of dredged material," Miscellaneous Paper D-94-4, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

1 Introduction

Sediment toxicity bioassays are often conducted to support the evaluation of dredged material proposed for open-water disposal under Section 103 of the Marine Protection, Research, and Sanctuaries Act (MPRSA) of 1972 (Public Law (PL) 92-532) and Section 404(b)(1) of the Federal Water Pollution Control Act of 1972 (PL 92-500), as amended by the Clean Water Act (CWA) of 1977 (PL 95-217). Bioassays are conducted to evaluate the potential toxicity of sediment-associated contaminants such as heavy metals, petroleum hydrocarbons, and chlorinated organics. Developing these bioassays requires considerable time and research in a variety of topic areas. Some bioassays are intuitively more developed and more appropriate for regulatory evaluations than others. Judging the developmental status of sediment toxicity bioassays for the regulatory evaluation of dredged material has been difficult because of a lack of developmental criteria. The paradigm described herein helps meet that need.

This report describes a paradigm for developing sediment toxicity bioassays for the regulatory evaluation of dredged material. This developmental paradigm serves several functions. It provides a framework for judging the developmental status of any sediment bioassay. This permits both scientific and regulating personnel to gauge a test's completeness any time during its evolution. It also allows one to compare the relative development of multiple bioassays. Describing the general pattern of test development will readily reveal gaps in our knowledge. This will permit investigators and program managers to optimize limited resources by directing research to areas needing immediate attention. Describing a logical sequence for test development should accelerate the development of sediment bioassays still in the conceptual stage. Finally, the latter stages of this paradigm suggest a process whereby the U.S. Army Corps of Engineers (USACE) and the U.S. Environmental Protection Agency (USEPA) can incorporate sediment toxicity bioassays into the evaluation of dredged material proposed for open-water disposal. This paradigm, therefore, serves the important function of interfacing science and public policy.

2 Approach

Initial Guidance (1992)

Written guidance for judging the developmental status of sediment toxicity bioassays did not exist when this project began (1992). Consequently, initial input was obtained by telephone from about 40 individuals in the scientific and regulatory communities. Each person was asked to describe the characteristics they would expect to see in a fully developed sediment toxicity bioassay for the regulatory evaluation of dredged material. Their input formed the basis for initial guidance for what constitutes a fully developed sediment toxicity bioassay for the regulatory evaluation of dredged material (Dillon 1992).

Peer-Review Written Comments and Workshop (1993)

The following year, Dillon (1992) was extensively peer reviewed. Written comments were solicited and received from a broad group of individuals. Follow-up discussions took place at a workshop held 16-17 June 1993 in Denver, CO. The written comments and workshop discussions were used to formulate the final developmental paradigm for sediment toxicity bioassays reported herein. Over 70 individuals provided input during this 2-year project (Table 1).

3 Results

General Evolutionary Pattern of Sediment Toxicity Bioassays

The development of sediment toxicity bioassays usually progresses through a series of stages or phases. At each stage, the protocol is modified to reflect improvements demonstrated through research and experience. Initially, an investigator will propose a bioassay and conduct preliminary laboratory research. If the bioassay looks promising, other investigators with different perspectives and backgrounds may evaluate the bioassay. This independent research and development exposes both strengths and weaknesses of the proposed sediment bioassay. Tests that survive this peer scrutiny may become widely accepted in the scientific community. If so, succeeding activities concentrate less on test development and more on performance, e.g., the bioassay's discriminatory power. If a broad consensus develops and sufficient research has been conducted, a standard method may be published usually by an independent standard-setting organization; e.g., American National Standards Institute (ANSI), American Society for Testing and Materials (ASTM), American Society for Quality Control (ASQC), in *Standard Methods*. The standardized sediment bioassay test method may then be considered by the two Federal agencies that have statutory responsibility for the national dredging program (USACE and USEPA). These agencies evaluate the appropriateness of the bioassay from the perspective of managing a major regulatory program. This final phase represents the interface between science and public policy.

A Developmental Paradigm for Sediment Toxicity Bioassays

Based on input from a wide variety of individuals (see Table 1), a developmental paradigm for sediment toxicity bioassays was created (Table 2). This paradigm reflects the general evolutionary pattern of sediment bioassays described above, has been extensively peer reviewed, and represents broad scientific consensus. Elements of the developmental paradigm are described below.

Phase I - Initial development by test proponent

Phase I Scoping

Rationale. The test proponent must explain how the proposed sediment bioassay will be used in the regulatory evaluation of dredged material. This obviously requires some knowledge of the regulatory milieu. Without a clear rationale, considerable resources may be expended developing a test for which there is no practical use. For example, is the bioassay intended to evaluate deposited or suspended sediments? Is it designed for early tier screening or more detailed later evaluations? Will it be used in the ocean or inland disposal programs?

The rationale may describe other applications beyond dredged material testing (e.g., field surveys of in situ toxicity and risk assessments). Existing bioassays developed for other purposes must still include a rationale specific to the regulatory evaluation of dredged material.

Assessment/measurement end points. The concept of assessment and measurement end points was adopted by USEPA while developing its Framework for Ecological Risk Assessment (USEPA 1992). Assessment end points are "formal expressions of the actual environmental value that is to be protected" (Suter 1990). Assessment end points are identified by managers and reflect their decision-making environment (e.g., societal concerns). Examples might include the following: (a) safeguard local fish and shellfish populations, (b) maintain healthy benthic habitat, (c) protect threatened and endangered species, etc. A measurement end point is "a measurable environmental characteristic that is related to the valued characteristic chosen as the assessment endpoint" (Suter 1990). Measurement end points are identified by the scientific/technical communities. Ecological models are often used to link measurement end points to assessment end points. For example, if the manager wishes to protect a local fishery resource, a population demographic model may be used to link observed effects on individual organisms to future population viability. Developing assessment and measurement end points represents a dialog between the environmental manager and the scientific/technical community. All too often this dialog occurs only after the technical work has been completed.

The concept of assessment and measurement end points is used in the USEPA/USACE dredging program and has direct application to the development of sediment toxicity bioassays. Federal statutes (MPRSA and CWA) require "no unacceptable adverse impacts" on the environment as a result of dredged material disposal. This statutory language is a national assessment end point. Several measurement end points are used to evaluate this assessment end point. One of these, sediment toxicity, has traditionally been measured by determining survival of very sensitive test species after short-term exposures. Selection of test end points and test species, therefore, are two of the first and most important steps in the development of sediment toxicity bioassays.

Test end points. Dredged material toxicity bioassays have relied heavily on survival as a test end point. While this will continue to be true, a new generation of sediment bioassays with sublethal end points and longer (chronic) sediment exposures is now evolving (Dillon 1993). Although the potential number of sublethal end points is very large, the practical number of usable end points for regulatory programs is much smaller. Test end points must be ecologically relevant, easily understood by the general public, and not too difficult to routinely measure. End points must also harmonize with the national assessment end point of "no unacceptable adverse impacts." Reproduction and growth are two highly desirable sublethal test end points that meet these criteria and enjoy broad scientific consensus (Dillon, Gibson, and Moore 1990).

Test species. Selection of the test species is critical to the success of the bioassay. Following are some important selection criteria.

a. *Available throughout the year.* Sufficient numbers of healthy organisms must be readily available throughout the year either through laboratory cultures or field collections.

b. *Handleable.* One must be able to routinely maintain and manipulate the test species (including shipping) as required by the laboratory protocol. Consistent, acceptable responses in the negative and positive controls must be achievable by contract laboratories.

c. *Compatible with test media.* The habitat, substrate, and nutritional requirements of the test species must harmonize with the test media. For example, infaunal species should be used to evaluate bedded sediments, while epibenthic, planktonic, or nektonic species are used with suspended sediments. Grain-size tolerance should be compatible with the test material.

d. *Appropriate sensitivity.* The test species must not be insensitive to major classes of environmental contaminants.

e. *Ecologically important.* The biology and natural history of test species should document its ecological importance; e.g., carbon flow and nutrient cycling.

f. *Commercially/recreationally important or indigenous.* Regional concerns to maintain and protect local populations of biological resources may be an important consideration in species selection.

Literature review. All pertinent information should be documented by the test proponent in a thorough review and analysis of the literature.

Alpha protocol. Development of sediment toxicity bioassays is not unlike the development of computer software. A succession of "new and improved" versions appear with continual "debugging," refinements, and an expanding

user group. This paradigm uses similar terminology to connote progressive development of sediment bioassays. For example, *Phase I Scoping* ends with a suggested test method called the Alpha protocol. It is the first time the suggested bioassay procedure is put to paper. The Alpha protocol may or may not be published.

Phase I Laboratory research and development

Statistical design. Statistical design is the a priori description of what types of data and analytical methods are required to adequately test a given hypothesis. Rigorous experimental research begins with a sound statistical design. Important components include but are not limited to the following:

a. Hypothesis formulation.

b. Hypothesis testing (data reduction/data analysis).

c. Level of significance (α error).

d. Power analysis (β error).

e. Number of treatments, number of replicates/treatment.

f. Intralab precision.

Experimental design and procedure. Experimental design is a detailed description of how the statistical design will be implemented and the bioassay conducted. It includes but is not limited to the following:

a. Manipulation of sediment before, during, and after the test.

b. Manipulation of test species before, during, and after the test.

c. Physical conditions (temperature, photoperiod, aeration, etc.).

d. Replicate description (size, number of animals/replicate, etc.).

e. Feeding regime.

f. Daily activities (visual observations, water quality, etc.).

g. Duration of test.

h. Test initiation/termination procedures.

Quality assurance (QA)/quality control (QC). QA/QC are the administrative and technical steps taken to ensure reliable data are generated with

specified precision and accuracy. They include but are not limited to the following:

a. Good laboratory practices (GLPs).

b. Standard operating procedures (SOPs).

c. Acceptable response in negative controls.

d. Consistent response in positive controls/use of control charts.

e. Data audits.

f. Corrective action procedures.

Test ruggedness. American Society of Testing and Materials (ASTM) (1992a) defines "ruggedness" as the "insensitivity of a test method to departures from specified test or environmental conditions." Some "departures" can be managed through good laboratory practices, a well-developed QA/QC program, and by strict adherence to a published standard protocol. Other aspects of test ruggedness, however, are more problematic. Nontreatment factors not related to sediment-associated contaminants can originate from three sources: (a) geophysical properties of dredged material, (b) health and condition of the test species, and (c) unforeseen deviations in experimental conditions. Some potentially important nontreatment factors include the following:

a. Grain size.

b. Ammonia/sulfide toxicity.

c. Interstitial salinity/hardness.

d. Macronutrients and micronutrients.

e. Sediment manipulation.

f. Feeding regime.

g. Water movement (static, static-renewal, flow-through).

h. Seasonal/reproductive condition.

i. Acclimation.

j. Presence of indigenous predators/competitors.

k. Shipping.

Nontreatment factors can bias test results of acute lethality sediment bioassays (DeWitt, Ditsworth, and Swartz 1988; Ankley, Katko, and Arthur 1990; Word et al. 1991). Their potential influence will increase when chronic sublethal sediment bioassays are more widely used. It is important, therefore, to address nontreatment factors during test development. Guidance for experimental determination of test ruggedness is available (ASTM 1992b). Ruggedness may also be evaluated by testing sediments representing a wide range of nontreatment factors. Whatever the approach, results should be summarized as a matrix of conditions for which the sediment bioassay is or is not appropriate.

Dredged material testing. During Phase I, the bioassay should be conducted with samples of dredged material. Sediments should represent a broad spectrum of suspected toxicities and geophysical characteristics. Success (or failure) of this dredged material testing will be a function of the quality and quantity of preceding research and development.

Peer-reviewed publications. The test proponent must communicate research results in the peer-reviewed literature. These publications serve several functions. First, they permit simultaneous access to the test protocol. Prior to publication, knowledge of a particular bioassay is anecdotal and generally limited to informal communications among colleagues. Acceptance for peer-review publication, however, does not necessarily imply broad acceptance by the scientific community.

Second, scrutiny during the peer-review process greatly increases the chances that weaknesses in the test method will be uncovered. This is a healthy process. Exposing weaknesses does not necessarily disqualify a test. On the contrary, it usually leads to significant improvements. At the very least, this scrutiny helps define the limits of the bioassay.

Third, authors of a good, well-written journal article will identify knowledge gaps and recommend important areas for further research and development. This discussion promotes scientific debate and stimulates other researchers. At this point, the sediment bioassay is beginning to move out of its laboratory of origin and into the larger family of research laboratories.

Beta protocol. After initial research results have been published, the preliminary Alpha protocol may be revised to a Beta version. The Beta version may be published as a methods journal article, a technical report, or in some other format. It may be prudent to provide informational copies to the USACE and USEPA in anticipation of eventual regulatory use. The published Beta protocol is what other laboratories follow in Phase II.

Phase II - Evaluation by multiple laboratories

Continued research and development. The research and development described in Phase I is rarely completed by the initial test proponent.

Evaluation by multiple laboratories greatly leverages the research effort by providing additional resources.

Interlaboratory studies. Interlaboratory studies are generally conducted to determine the statistical variability (precision) among laboratories. Factors contributing to this variation include (a) experience and expertise of the operator, (b) instrument type and calibration, and (c) the environment in which the test is conducted. ASTM (1992c) has provided a standard practice for conducting interlaboratory studies. For sediment toxicity bioassays, interlaboratory studies require a well-written test method that can be executed by participating laboratories. Great care must be taken to ensure each laboratory is testing the same dredged material at the same time. Successful interlaboratory studies demand considerable resources, committed participants, and proactive project management.

Interpretive guidance. Interpretive guidance is the technical information that regulating agencies use in judging the importance of bioassay results. Interpretive guidance may be based on laboratory experiments, field studies, and best professional judgment. For example, if the bioassay end points are survival, growth, and/or reproduction, the technical basis might be a calibrated population demographic model. Field studies involving contaminant gradients can also provide helpful insights for interpreting test results.

Because interpretive guidance can have several sources, it usually emerges only after multiple laboratories have evaluated the sediment bioassay. Interpretive guidance has both a statistical and biological component. The former includes the power of the test, intralaboratory and interlaboratory variability, etc. The latter refers to the biological significance of observed results. For example, if a particular dredged material reduces survival by 5 percent relative to the reference sediment, is that biologically significant? Would a 10-percent decrease be twice as bad or only incrementally injurious? Would a 50-percent reduction represent an order of magnitude increase in toxicity? Providing technically sound interpretive guidance for these and other test end points represents a significant challenge to the scientific community (Dillon 1993).

Testing with a wider range of dredged material. As multiple laboratories evaluate the sediment bioassay, the number of dredged materials tested will increase. These additional samples should represent a range of toxicities and geophysical characteristics.

Species sensitivity to major contaminants. Sensitivity of the test species to major classes of contaminants (metals, chlorinated hydrocarbons, petroleum hydrocarbons, and pesticides) should be documented. This information can aid in species selection and test interpretation. For example, if a particular sediment is contaminated with heavy metals, one might wish to select a test species that is especially sensitive to that class of contaminant. Relative contaminant sensitivity may help explain why some sediments are toxic to a particular test species, while others sediments are not. Likewise, it may

explain why exposure to one sediment resulted in significant mortalities in one species but not another.

Contaminant sensitivity information can be especially useful when expressed as a dose-response relationship. Klassen (1986) described this relationship as "the most fundamental and pervasive concept in toxicology." Its centrality is due to the fact that it establishes chemical-specific causality. Dose-response data identify what levels of a contaminant are toxic and, importantly, which levels are not toxic. Sediments are complex mixtures. Some knowledge of chemical-specific causality may be insightful or even essential in developing technically sound sediment bioassays. In field validation studies, causal dose-response data can corroborate correlative field responses observed along a contaminant gradient (see **Verification/Validation**). When combined with a knowledge of feeding behavior and microhabitat exposures, chemical-specific dose-response data can help distinguish sensitivity to contaminated sediments from contaminant sensitivity. Relative dose-response data are also especially powerful for assessing interspecific differences in xenobiotic metabolism for major contaminants.

Cost and logistics. By this point, the technical community should be able to document the cost and logistics associated with conducting the bioassay. Does it require extensive capitol outlay? Is it cheap and easy to run? What parts of the test are most difficult? Which procedures require intensive mentoring.

Peer-reviewed publications. As in Phase I, technical results generated in the multiple laboratories must be communicated in peer-reviewed publications.

Acceptance by the scientific community. Scientific acceptance of a sediment toxicity bioassay is a primary consideration of the user community. There is no written guidance for determining when scientific acceptance has been achieved. Rather, a "survival of the fittest" process usually takes place. After multiple laboratory evaluation, some bioassays are utilized with greater frequency, while others receive less and less attention. Some disappear from use altogether. This is a slow but healthy process. Close scrutiny by many investigators helps ensure survival of tests that work and are biologically meaningful. If this process has one weakness, it is determining when a particular test has been accepted (or rejected) by the scientific community. Based on discussions at the peer-review workshop in Denver, CO, the following criteria for scientific acceptance of a sediment toxicity bioassay were agreed upon.

a. Written protocol available.

b. Technical basis for protocol published in peer-reviewed journals.

c. Consistently used by multiple laboratories.

d. Provides interpretable results of environmental significance.

Gamma protocol. Phase II terminates with a Gamma version of the sediment bioassay. This upgrade of the Beta version incorporates all experiences and research results of multiple laboratories. Again, sending informational copies to the regulating agencies may be prudent. In some instances, the Beta or Gamma protocol may be an ASTM standard guide produced by Subcommittee E47.03 on Sediment Toxicity. ASTM (1992d) defines a standard guide as "a series of options or instructions that do not recommend a specific course of action."

By definition, ASTM guides are not step-by-step "cookbooks." For that, ASTM has another end product, the standard test method, which they define as "a definitive procedure for the identification, measurement, and evaluation of one or more qualities, characteristics, or properties of a material, product, system or service that produces a test result."

Phase III - Development of a standard test method

Intertest comparisons. Intertest comparisons evaluate the sediment bioassay's discriminatory power. That is, how frequently and with what precision does the test indicate toxicity relative to other sediment bioassays. Most intertest studies have found that no single bioassay is consistently the most sensitive and precise (Burton et al. 1989; Giesy and Hoke 1989; Long and Buckman 1989; Pastorok and Becker 1990). For that reason, a frequent recommendation is to use a battery of sediment bioassays. To be valid, intertest studies should be conducted on the same sediment, at the same time, and, ideally, side by side in the same laboratory. Unique features of the bioassay (e.g., recommended temperature and salinity) should be retained.

Verification/validation. There are no formal guidelines for verifying/validating sediment toxicity bioassays for dredged material testing. Comparable guidance, however, can be gleaned from the field of ecological modeling (Jorgensen 1988). An ecological model is said to be validated when model outputs approximate real world values. That is, the model is predictive. Validation occurs near the end of model development, after equation calibration, sensitivity analysis, and model verification. Data for model validation must be collected independently from data used to calibrate the model. In addition, the domain of validation data should represent a wide range of forcing functions.

Verification, on the other hand, occurs early in model development, often iteratively with model calibration and sensitivity analysis. Verification is an evaluation of the model's behavior and internal logic. An ecological model is said to be verified when "it behaves in the way the model builder wanted it to behave" (Jorgensen 1988). Obviously, verification is a very subjective process. In contrast, validation is the objective independent evaluation of the model's predictive capability.

In the context of dredged material testing, sediment toxicity bioassays are simplistic ecological models. As measurement end points, they quantitate potential changes in the Federal assessment end point of "no unacceptable adverse impacts" (see **Assessment/measurement end points**). Verification of these ecological models occurs when the sediment bioassay behaves as expected. In contrast, validation of a sediment toxicity bioassay has been achieved only when it has been shown to be predictive of dredged material disposal impacts. In the context of ecological models, verification and validation are distinctly different terms requiring disparate levels of effort to achieve.

Verification of sediment toxicity bioassays may be accomplished in at least two ways. One is to conduct the bioassay with samples of dredged material representing a range of suspected toxicities. If it responds as expected, it is verified. On a smaller scale, one can dilute a toxic sediment and expect to see a corresponding response gradient in bioassay results. In either case, the bioassay's state variables (i.e., the test end points) are responding as expected to increasing forcing functions (i.e., the sediment-associated contaminants). Many sediment toxicity bioassays have been verified.

In contrast, few sediment bioassays have been validated. To do so, one first makes a prediction regarding dredged material impacts. This prediction is based on predisposal bioassays. One then monitors the disposal event to see if the model predictions for those sediments were correct. Field validation has rarely been attempted for large-scale dredging projects. A more feasible approach would utilize mesocosms or small-scale disposal events. Here, validation of the bioassay prediction for a variety of materials could be assessed with precision and accuracy. Bioassay response to sediments collected along a natural pollution gradient may also provide valuable insights. However, since these are correlative observations, one must be able to exclude the influence of covarying parameters and historical events unrelated to present levels of sediment contamination. One way to strengthen these correlative data is to corroborate with chemical-specific causal relationships generated in the laboratory (see **Species sensitivity to major contaminants**). Clearly, further discussions on the verification and validation of sediment toxicity bioassays for dredged material testing are warranted.

Peer-reviewed publications. As in previous phases, results must be communicated in peer-reviewed publications.

Protocol published by standard-setting group. Phase III ends with the publication of the sediment bioassay as a standard method. Possible outlets include ANSI, ASTM, ASQC, *Standard Methods*, etc.

Phase IV - Evaluation by user groups

Once a standard method for the sediment bioassay exists, user groups can evaluate its utility for the regulatory program. Users include the following: (a) the Federal regulating agencies (USACE and USEPA), (b) the States in the

case of Section 404(b)(1) evaluations, (c) the bioassay contracting community, and (d) members of the private sector seeking permits to dredge. Phase IV represents the interface between science/technology and public policy. This interfacing occurs on both a national and regional basis.

Joint agency consideration (USEPA/USACE). Responsibility for the dredging program is shared jointly by the USEPA and USACE. The USACE evaluates the material, and the USEPA reviews the evaluation. For Section 103 actions, bioassays must be jointly approved by both agencies. Their evaluation of a candidate sediment bioassay begins with its scientific merits, but includes other less technical issues. As public servants and custodians of Federal monies, the USACE and USEPA are required to consider and balance resource expenditures with benefits received for all Federal actions. They must be able to explain to the public or, in the case of permitted activities, to the private sector precisely why the sediment bioassay is being conducted, what information it will yield, and how that information will contribute to decision making. Important considerations include but are not limited to the following:

a. Relevant and appropriate for the intended use.

b. Founded in the applicable laws and regulations.

c. Accepted by the scientific community.

d. Accompanied by sound interpretive guidance.

e. Demonstrated track record with a range of dredged material.

f. Cost-effective.

g. Simplified "cookbook" available.

h. "Doable" in a routine fashion by contract laboratories.

i. Able to sustain judicial review.

Most sediment bioassays reaching this stage will likely satisfy most agency concerns and considerations. If not, the agencies may recommend that the bioassay be "recycled" to an earlier phase for further research and development.

Training with instructional "cookbook." There is an initial "learning curve" whenever a facility attempts a new test. This is true for all laboratories, experienced or novice, and for all sediment bioassays, simple or complex. It is highly desirable that this learning curve plateau before bioassays are conducted for regulatory decision making. This requires some form of training. Training may be in the form of mentoring where a novice investigator learns from one more experienced. Training may also be provided by the regulating

agencies. Alternatively, the private sector may find it profitable to offer this training as a service.

Whatever the mechanism, this training should be carried out using a simplified step-by-step instructional cookbook. This cookbook should be based on the standard test method. However, including all the detailed supporting documentation may not be necessary to conduct the bioassay in a technically sound manner. The cookbook should include standardized formats for recording data, QA/QC, reporting test results, etc. Training with a cookbook will help establish market-based costs for the bioassay and evaluate the contract laboratory's ability to conduct the test. An important measure of laboratory success during the training period is routinely meeting or exceeding the performance criteria for negative and positive controls.

Joint agency recommendations. Once training is complete, the USEPA/USACE may recommend the following: (a) incorporation of the sediment toxicity bioassay into the regulatory program, (b) conditional use based on the performance of additional research, or (c) further research and development prior to regulatory use. The clearest sign of joint agency approval is inclusion in Section 103 and/or Section 404 Implementation Manuals.

Periodic review. Once incorporated in the regulatory program, performance of the sediment bioassay should be reviewed periodically. This review will indicate if the test is performing as expected and reveal any unanticipated problems. Periodic reviews serve as forums for technology transfer, allowing users and the scientific community the opportunity to share experiences and exchange information.

4 Peer-Review Comments

The developmental paradigm for sediment toxicity bioassays reported here reflects extensive peer-review comments and discussions. It thus represents broad scientific consensus. In addition to general comments, reviewers were asked to critique three specific aspects of the paradigm: completeness, sequencing, and relative importance of developmental activities. Below is a summary of their comments.

Completeness of the Developmental Paradigm

All reviewers indicated that the paradigm was complete. A few reviewers even thought it was too complete. That is, most of the sediment bioassays in use today have not been developed to the extent suggested by the paradigm. This is true. All sediment bioassays lie on a developmental continuum. Some are more complete than others. This paradigm describes the developmental process as it should occur. It was never intended to set pass-fail criteria for specific bioassays.

Sequencing of Developmental Activities

All reviewers indicated that the sequencing of developmental activities was good. A few were concerned that the paradigm gave the false impression that research and development proceeds sequentially in a fixed linear fashion. Practitioners realize this is not true. Rather, the process is often iterative with some aspects of research and development proceeding in a sequence not consistent with that described in the paradigm for others. Some activities, in fact, are accomplished simultaneously. Still, some efforts logically precede others (e.g., a written protocol before interlaboratory studies). The paradigm was not meant to advocate strict conformity to a particular sequence. It does, however, describe a logical progression for developing sediment toxicity bioassays.

Relative Importance of Developmental Activities

Peer reviewers indicated that all the developmental activities in the paradigm were important. Given that consensus, the question then becomes, to

what extent must each be addressed? For example, how much information is necessary to address the issue of test ruggedness? How many and what kinds of sediment satisfy the requirement "Testing with a Wider Range of Dredged Material"? What types of data and study designs are necessary to field verify or field validate a sediment toxicity bioassay? These questions have no simple answers? The type and amount of information to address each developmental activity will have to be made on a case-by-case basis in the context of the individual sediment bioassay.

5 Summary

a. A paradigm for developing sediment toxicity bioassays for the regulatory evaluation of dredged material has been established.

b. This paradigm is phased and parallels the evolution of many sediment toxicity bioassays; i.e., initial development by the test proponent, peer scrutiny by multiple laboratories, consensus on a standard test method, and incorporation into the USEPA/USACE regulatory dredging program.

c. The developmental paradigm has been extensively peer reviewed and reflects the input from over 70 scientists and regulators.

d. Peer-review comments indicate the paradigm is complete and follows a logical sequence. Sediment bioassays that have addressed each element of the paradigm may be considered developed. However, the amount and type of information required for each developmental activity varies on a case-by-case basis for each sediment bioassay.

e. Scientific acceptance of a sediment toxicity bioassay has occurred when (a) written protocol is available, (b) the technical basis for the protocol has been published in peer-reviewed journals, (c) the bioassay is consistently used by multiple laboratories, and (d) the bioassay provides interpretable results of environmental significance.

f. Technically sound interpretive guidance, especially for chronic sublethal sediment bioassays, remains a significant challenge to the scientific community.

g. Clear guidance for what constitutes field validation and/or verification of sediment toxicity bioassays is needed.

h. A period of training is critical if sediment bioassays are to be successfully conducted on a routine basis. At the present time, this training is accomplished through informal mentoring arrangements. Institutionalized training may reduce the frequency of invalid and/or suspect bioassay results.

i. Although designed specifically for dredged material toxicity bioassays, this paradigm can be adapted for developing sediment toxicity bioassays for many other applications; e.g., bioaccumulation, field surveys of in situ toxicity, environmental risk assessments, etc.

References

American Society of Testing and Materials. (1992a). "Standard terminology relating to quality and statistics," Annual Book of Standards, Philadelphia, PA, Vol. 14.02, 371-377.

_____. (1992b). "Standard guide for conducting ruggedness tests," Annual Book of Standards, Philadelphia, PA, Vol. 14.02, 744-749.

_____. (1992c). "Standard practice for conducting an interlaboratory study to determine the precision of a test method," E 691-92 Annual Book of Standards, Philadelphia, PA, Vol. 14.02, 491-510.

_____. (1992d). Annual Book of Standards, Philadelphia, PA, Vol. 14.02, iii.

Ankley, G. T., Katko, A., and Arthur, J. W. (1990). "Identification of ammonia as an important sediment-associated toxicant in the lower Fox River and Green Bay, Wisconsin," *Environ. Toxicol. Chem.* 9, 313-322.

Burton, G. A., Jr., Stemmer, B. L., Winks, K. L., Ross, P. E., and Burnett, L. C. (1989). "A multitrophic level evaluation of sediment toxicity in Waukegan and Indiana Harbors," *Environ. Toxicol. Chem.* 8, 1057-1066.

DeWitt, T. H., Ditsworth, G. R., and Swartz, R. C. (1988). "Effects of natural sediment features on survival of the phoxocephalid amphipod *Rhepoxynius abronius*," *Mar. Environ. Res.* 25, 99-124.

Dillon, T. M. (1992). "Determining the developmental status of sediment toxicity bioassays," Technical Note EL-EEDP-01-29, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

_____. (1993). "Developing chronic sublethal sediment bioassays: A challenge to the scientific community. *Toxicology and risk assessment: 2nd Volume, ASTM STP 1173.* J. W. Gorsuch, F. J. Dwyer, C. G. Ingersoll, and T. W. La Point, ed., American Society for Testing and Materials, Philadelphia, PA.

Dillon, T. M., Gibson, A. B., and Moore, D. M. (1990). "Development of chronic sublethal sediment bioassays: Proceedings of a workshop," Technical Note EL-EEDP-01-22, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Giesy, J. P., and Hoke, R. A. (1989). "Freshwater sediment toxicity bioassessment: Rationale for species selection and test design," *J. Great Lakes Res.* 15, 539-569.

Jorgensen, S. E. (1988). *Fundamentals of ecological modeling.* Elsevier, NY.

Klassen, C. D. (1986). "Principles of toxicology." *Casarett and Doull's Toxicology.* C. D. Klassen, M. O. Amdur, and J. Doull, ed., 3rd ed., Macmillan Publishing Co., New York, 18.

Long, E. R., and Buchman, M. F. (1989). An evaluation of candidate measures of biological effects for the National Status and Trends Program. NOAA Technical Memorandum NOS OMA 45, Seattle, WA.

Pastorok, R. A., and Becker, D. S. (1990). "Comparative sensitivity of sediment toxicity bioassays at three Superfund sites in Puget Sound." *Aquatic toxicology and risk assessment: Thirteenth Volume, ASTM STP 1096.* W. G. Landis and W. H. van der Schalie, ed., American Society for Testing and Materials, Philadelphia, PA, 123-139.

Suter, G. W., II. (1990). "Endpoints for regional ecological risk assessments," *Environmental Management* 14, 9-23.

U.S. Environmental Protection Agency. (1992). "Framework for ecological risk assessment," EPA/630/R-92/001.

Word, J. W., Claiborne, B. W., Ward, J. A., and Chapin, C. (1991). "The effect of test sediment stabilization and disturbance on acute toxicity to the amphipod *Rhepoxynius abronius.*" *Proceedings of Puget Sound Research '91, 4-5 Jan 91, Seattle, WA.* Puget Sound Water Quality Authority, Olympia, WA, 441-448.

Table 1
Individuals Contributing to the Developmental Paradigm

Adams, W. ¶†§	ABC Laboratories, Columbia, MO
Alden, R. ¶	Old Dominion University, Norfolk, VA
Anderson, D. ¶	USACE, St. Paul District, St. Paul, MN
Ankley, G. ¶	USEPA, Environmental Research Laboratory, Duluth, MN
Armitage, T. †§	USEPA, Headquarters, Office of Water, Washington, DC
Arnold, R. §	Exxon Biomedical Sciences Inc., East Millstone, NJ
Barton, J. ¶	USEPA, Region X, Seattle, WA
Bay, S. ¶†	So. CA Coastal Water Research Project, Long Beach, CA
Black, J. †	EA Engineering, Science and Technology, Sparks, MD
Bridges, T. §	USACE, Waterways Experiment Station, Vicksburg, MS
Burton, G. ¶†§	Wright State University, Dayton, OH
Call, D. ¶†	University of Wisconsin-Superior, Superior, WI
Cardwell, R. †	Parametrix, Seattle, WA
Casillas, E. ¶†§	NOAA, National Marine Fisheries Service, Seattle, WA
Chapman, P. ¶†§	EVS Consultants, Ltd., North Vancouver, BC, Canada
Chase, T. §	USEPA, Headquarters, Office of Water, Washington, DC
Cowgill, D. ¶	USEPA, Great Lakes National Program Office, Chicago, IL
Day, K. †§	Environment Canada, Burlington, Ontario, Canada
Dinnell, P. ¶†	University of Washington, Seattle, WA
Dorkin, J. ¶	USEPA, Region V, Chicago, IL
Fong, C. §	USACE, San Francisco District, San Francisco, CA
Fredette, T. ¶†	USACE, New England Division, Waltham, MA
Fritz, A. †	NOAA, HAZMAT Response and Assessment Division, Seattle, WA
Guy, K. §	USACE, San Francisco District, San Francisco, CA
Hall, J. ¶	Texaco, Inc., Port Arthur, TX
Hansen, D. ¶	USEPA, Environmental Research Laboratory, Narragansett, RI
Hollar, K. ¶	USEPA, Region VI, Dallas, TX
Ingersoll, C. ¶†§	U.S. Fish and Wildlife Service, Columbia, MO
Johnson, D. §	USEPA, Region IV, Atlanta, GA

(Sheet 1 of 3)

Note:
¶ = Provided initial verbal input in 1992.
† = Provided written comments in 1993 on Dillon (1992).
§ = Workshop participant in Denver, CO, 16-17 Jun 93.

Table 1 (Continued)

Kendall, D. ¶†	USACE, Seattle District, Seattle, WA
Klein, L. §	USACE, New York District, New York, NY
Lamberson, J. ¶†§	USEPA, Environmental Research Laboratory, Newport, OR
Landrum, P. †§	NOAA, Great Lakes Environmental Research Lab, Ann Arbor, MI
Lazorchak, J. ¶	USEPA, Cincinnati, OH
Lemlich, S. ¶	USACE, Waterways Experiment Station, Vicksburg, MS
Lutz, C. §	USACE, Waterways Experiment Station, Vicksburg, MS
MacDonald, D. †	NOAA, HAZMAT Response and Assessment Division, Seattle, WA
Malek, J. §	USEPA, Region X, Seattle, WA
Mathies, L. ¶†	USACE, New Orleans District, New Orleans, LA
Matta, M. †	NOAA, HAZMAT Response and Assessment Division, Seattle, WA
Miller, J. ¶	USACE, North Central Division, Chicago, IL
Moore, D. §	USACE, Waterways Experiment Station, Vicksburg, MS
Nacci, D. ¶	Science Applications International Corp., Narragansett, RI
Nelson, M. ¶†	U.S. Fish and Wildlife Service, Columbia, MO
Nelson, W. ¶†	USEPA, Environmental Research Laboratory, Narragansett, RI
Norberg-King, T. §	USEPA, Environmental Research Laboratory, Duluth, MN
Oshida, P. ¶	USEPA, Region IX, San Francisco, CA
Pennington, R. ¶	USACE, Jacksonville District, Jacksonville, FL
Peltier, W. ¶	USEPA, Region IV, Athens, GA
Pickard, J. §	USACE, Buffalo District, Buffalo, NY
Redlinger, J. §	USACE, Portland Division, Portland, OR
Rees, S. ¶	USACE, Mobile District, Mobile, AL
Reese, J. ¶†	USACE, North Pacific Division, Portland, OR
Rodriguez, G. §	USEPA, Region VIII, Denver, CO
Rosman, L. §	USACE, New York District, New York, NY
Ross, B. ¶†	USEPA, Region IX, San Francisco, CA
Rubinstein, N. ¶§	USEPA, Environmental Research Laboratory, Narragansett, RI
Scott, K. ¶§	Science Applications International Corp., Narragansett, RI
Southerland, B. §	USEPA, Headquarters, Office of Water, Washington, DC
Spies, R. †	Applied Marine Science, Livermore, CA
Swartz, R. †§	USEPA, Environmental Research Laboratory, Newport, OR
Tabatabai, F. §	USACE, Los Angeles District, Los Angeles, CA

Table 1 (Concluded)

Tavolaro, J. ¶†	USACE, New York District, New York, NY
Tuchman, M. ¶	USEPA, Region V, Chicago, IL
Urabeck, F. ¶	USACE, Seattle District, Seattle, WA
Ward, J. §	Battelle Northwest Pacific Laboratory, Sequim, WA
Weber, C. ¶	USEPA, Cincinnati, OH
Welch, T. §	USACE, Los Angeles District, Los Angeles, CA
Wilson, J. §	USACE, Headquarters, Washington, DC
Word, J. ¶†§	Battelle Northwest Pacific Laboratory, Sequim, WA
Zoutendyk, D. §	USACE, Los Angeles District, San Diego, CA

Table 2
A Paradigm for Developing Sediment Toxicity Bioassays for the Regulatory Evaluation of Dredged Material

Phase I - Initial Development by Test Proponent
Phase I Scoping
· Rationale
· Assessment/measurement end points
· Test end points
· Test species
· Literature review
· Alpha protocol
Phase I Laboratory Research and Development
· Statistical design
· Experimental design and procedure
· QA/QC
· Test ruggedness
· Dredged material testing
· Peer-reviewed publications
· Beta protocol
Phase II - Evaluation by Multiple Laboratories
· Continued research and development
· Interlaboratory studies
· Interpretive guidance
· Testing with a wider range of dredged material
· Species sensitivity to major contaminants
· Cost and logistics
· Peer-reviewed publications
· Acceptance by the scientific community
· Gamma protocol

(Continued)

Table 2 (Concluded)
Phase III - Development of a Standard Test Method
· Intertest comparisons
· Verification/validation
· Peer-reviewed publications
· Protocol published by standard-setting group
Phase IV - Evaluation by User Groups
· Joint agency consideration (USEPA/USACE)
· Training with instructional cookbook
· Joint agency recommendations
· Periodic review

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1994	Final report

4. TITLE AND SUBTITLE
A Paradigm for Developing Sediment Toxicity Bioassays for the Regulatory Evaluation of Dredged Material

5. FUNDING NUMBERS

6. AUTHOR(S)
Thomas M. Dillon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Waterways Experiment Station
3909 Halls Ferry Road, Vicksburg, MS 39180-6199

8. PERFORMING ORGANIZATION REPORT NUMBER
Miscellaneous Paper D-94-4

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Corps of Engineers, Washington, DC 20314-1000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Available from National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Sediment toxicity bioassays may be required in the regulatory evaluation of dredged material. Use of these tests in a major regulatory program implies they have been fully developed. This is not always true. Some bioassays are intuitively more appropriate and more advanced than others. Judging their relative developmental status, however, has been difficult because of a lack of guidance. To meet this need, input from users and developers of sediment bioassays was solicited over a 2-year period. That input is summarized in a four-phased developmental paradigm. Phase I includes activities often conducted by the initial test proponent; e.g., species/end point selection, statistical/experimental design, and quality assurance/quality control. Phase II involves evaluation of the bioassay by multiple laboratories. Phase II concludes with acceptance of the sediment bioassay by the scientific community. Activities under Phase III include intertest comparisons and field validation/verification. Phase III terminates with the publication of a standard method by some standard-setting organization such as the American Society of Testing and Materials. Phase IV recommends a process for regulatory agencies (U.S. Environmental Protection Agency/U.S. Army Corps of Engineers) to judge the appropriateness of bioassays vis-a-vis dredged material evaluations.

14. SUBJECT TERMS
Bioassays
Development
Sediment

15. NUMBER OF PAGES
28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED		